DTIF DIGITAL TRUST & INTEROPERABILITY FRAMEWORK

Enabling Trustworthy Autonomous Science & Discovery Economy

SELF-DRIVING LABORATORIES (SDL)

- L0 Manual
- L1 Assisted
- L2 Automated
- L3 Adaptive
- L4 Collaborative
- L5 Sovereign

AI Hypothesis Engine

DIGITAL TRUST SCORE (DTS)

- Verifiability
- Observability
- Traceability
- Uncertainty
- Reproducibility

DTIF Autonomous Science Ecosystem

FRYX INTEROPERABILITY PROTOCOL

Data · Models · Robots · Agents · Cloud

DISCOVERY ECONOMY

Knowledge → Innovation → Validation → Impact

Trust • Transparency • Collaboration • Sovereignty

The Digital Trust & Interoperability Framework (DTIF): A Universal Architecture for Autonomous Science and Sovereign Discovery Economies

Er. Nabal Kishore Pande

ABSTRACT

Artificial intelligence and robotic laboratory automation are reshaping the foundations of scientific discovery. While AI systems can generate experimental hypotheses at unprecedented scale, existing research infrastructures lack the interoperability, trust verification mechanisms, and governance frameworks required to support reliable autonomous experimentation. This paper introduces the **Digital Trust & Interoperability Framework (DTIF)**, a unified architecture for governing autonomous scientific ecosystems. DTIF integrates three components: the **Self-Driving Laboratory (SDL) Maturity Model**, which standardizes levels of laboratory autonomy; the **Digital Trust Score (DTS)**, a quantitative metric for evaluating the reliability and auditability of autonomous research agents; and the **FRYX Interoperability Protocol**, enabling coordinated multi-agent experimentation across heterogeneous laboratory systems. Together, these components establish an operational foundation for trustworthy autonomous science and the emerging **Discovery Economy**, in which distributed laboratories continuously generate validated knowledge and economically valuable discoveries.

Keywords

Autonomous Laboratories

Self-Driving Laboratories

Autonomous Scientific Discovery

Digital Trust in AI Systems

Research Interoperability Frameworks

Digital Trust Score (DTS)

FRYX Interoperability Protocol

AI Governance in Scientific Research

Scientific Sovereignty

Multi-Agent Research Systems

Autonomous Experimentation

Discovery Economy

AI-Driven Research Infrastructure

Table of Contents

1. Introduction — The Bottleneck in Autonomous Discovery 6
1.1 The Computational–Experimental Paradox 6
1.2 The Trust Gap in Autonomous Research 7
1.3 Scientific Sovereignty Challenges 7
1.4 Research Objective 8
2. The Self-Driving Laboratory (SDL) Maturity Model 10
2.1 Level 0 — Manual Research Systems 10
2.2 Level 1 — Automated Execution 11
2.3 Level 2 — AI-Assisted Research 11
2.4 Level 3 — Closed-Loop Discovery 12
2.5 Level 4 — Collaborative Multi-Agent Research 12
2.6 Level 5 — Sovereign Autonomous Discovery 12
3. Digital Trust Scoring (DTS) Framework 14
3.1 DTS Formula and Parameter Overview 14
3.2 Verifiability 15
3.3 Observability 15
3.4 Traceability 16
3.5 Model Uncertainty 16
3.6 Trust Thresholds 16
4. FRYX Interoperability Protocol 18
4.1 Physical Execution Layer 18
4.1.1 Instrument API Gateways 18
4.1.2 Robotic Experiment Controllers 19
4.1.3 Sensor Data Capture Systems 19
4.2 Logic and Intelligence Layer 19
4.2.1 Hypothesis Generation Engine 19
4.2.2 Universal Model Translator 19
4.2.3 Multi-Agent Coordination System 20
4.3 Governance and Trust Layer 20
4.3.1 Digital Trust Engine 20
4.3.2 Policy-as-Code Enforcement 20
4.3.3 Cross-Agent Validation Protocols 21
5. Autonomous Discovery Workflow 22
6. The Discovery Economy Model 24
6.1 From Cost Centers to Discovery Infrastructure 24

6.2 Regional Autonomous Research Hubs 24
6.3 Global Networks of Autonomous Research Nodes 25
7. Implications for Scientific Governance 27
7.1 Open Interoperability Standards 27
7.2 Transparent AI Governance 27
7.3 Global Verification Protocols 28
7.4 Ethical and Equity Considerations 28
8. Conclusion 30
Appendices 32
Appendix A — SDL Maturity Assessment Template 32
A.1 Assessment Scope and Method 32
A.2 Assessment Instrument Structure 32
A.3 SDL Level Certification 32
Appendix B — Digital Trust Score Calculation Guide 34
B.1 Verifiability Index Calculation 34
B.2 Observability Index Calculation 34
B.3 Traceability Constant Calculation 34
B.4 Model Uncertainty Quantification 35
Appendix C — FRYX Architecture Specification 36
C.1 Physical Execution Layer Specification 36
C.2 Logic and Intelligence Layer Specification 36
C.3 Governance and Trust Layer Specification 36
Appendix D — Cross-Agent Validation Protocol 38
D.1 Validation Request Specification 38
D.2 Validation Execution and Reporting 38
D.3 Consensus Evaluation and Certification 38

1. Introduction — The Bottleneck in Autonomous Discovery

The convergence of artificial intelligence and robotic laboratory automation is reshaping the foundations of scientific inquiry at a pace that prior generations of researchers could not have anticipated. For the first time in the history of science, computational systems can formulate, prioritize, and refine research hypotheses at a rate that far exceeds the capacity of human researchers to evaluate them experimentally. Simultaneously, robotic platforms capable of executing complex multi-step laboratory protocols with sub-millimetre precision and nanolitre-scale liquid handling accuracy have become commercially accessible. The intersection of these two technological trajectories — superintelligent hypothesis generation and robotic experimental execution — creates both an extraordinary opportunity and a formidable architectural challenge.

Despite the technical feasibility of autonomous scientific systems, no universally adopted framework exists to govern their operation, classify their capabilities, verify the trustworthiness of their outputs, or enable interoperability between competing implementations. This absence of foundational infrastructure represents the defining challenge of autonomous discovery. The Digital Trust & Interoperability Framework (DTIF) is proposed in this work as a response to that challenge: a unified architecture that provides a common language, a trust measurement system, and a communication protocol for autonomous research ecosystems.

This section delineates four foundational dimensions of the problem that DTIF addresses: the computational–experimental paradox, the trust gap in autonomous research, challenges of scientific sovereignty, and the research objective to which this framework responds.

1.1 The Computational–Experimental Paradox

Artificial intelligence systems — particularly large language models trained on scientific literature, protein structure databases, and chemical property repositories — can generate hundreds of thousands of candidate research hypotheses within a single computational session lasting hour. Generative molecular design models such as those deployed in pharmaceutical discovery can enumerate chemical structures with predicted biological activity across vast regions of chemical space, producing candidate libraries of a scale that would have required years of combinatorial chemistry to assemble under conventional research paradigms.

Physical laboratory infrastructure, however, remains constrained by throughput limitations that are fundamentally rooted in the physicochemical reality of experimental science. Even highly automated high-throughput screening facilities — representing the current state of the art in laboratory automation — are capable of validating on the order of tens to hundreds of experimental candidates per week under typical operating conditions. A representative instance of this structural disparity is illustrated below:

AI designs 100,000 molecular candidates

▼

Laboratory infrastructure validates only 10–100 candidates weekly

This disparity — spanning three to four orders of magnitude — creates a structural bottleneck that constrains the practical impact of AI-driven scientific discovery. The overwhelming majority of computationally generated hypotheses cannot be validated within any reasonable research timeline, creating a vast 'dark hypothesis space' of unexamined predictions that may contain high-value discoveries. Addressing this bottleneck requires not

merely faster robots, but architectural frameworks that intelligently prioritize, coordinate, and distribute experimental validation workloads across networks of autonomous laboratories. The DTIF provides the coordination layer necessary to make such networks function coherently.

Furthermore, the paradox is self-reinforcing: as AI models are trained on validated experimental data, more accurate hypotheses are generated, producing even larger candidate libraries. Without scalable validation infrastructure, the epistemic value of AI-generated hypotheses cannot be realized, and model improvement stagnates. The DTIF breaks this cycle by enabling distributed validation at network scale, feeding a continuous improvement loop across both experimental throughput and AI model accuracy.

1.2 The Trust Gap in Autonomous Research

The introduction of autonomous systems into the scientific process raises fundamental questions about the epistemic status of AI-generated research outputs. When a human researcher designs and executes an experiment, decades of institutional norms govern what constitutes an acceptable protocol, how results should be documented, and what standards of reproducibility are required for publication. These norms are imperfectly followed in practice — the reproducibility crisis in psychology, medicine, and materials science attests to this — but they provide a shared framework within which scientific claims can be evaluated, contested, and refined.

Autonomous research systems currently lack standardized mechanisms equivalent to these norms. Three dimensions of the trust gap are particularly critical:

- **Experiment verification** — There is no universal standard for confirming that an autonomous system executed its experimental protocol faithfully, that instruments were properly calibrated, or that environmental conditions were within acceptable bounds during data acquisition.
- **AI reasoning transparency** — The inference processes of deep learning models deployed in autonomous research agents are largely opaque. When an AI system recommends an experimental direction, the causal pathway from input data to recommendation is typically inaccessible to external evaluation.
- **Reproducibility tracking** — Autonomous systems produce data at rates that overwhelm conventional documentation practices. Without automated provenance systems capable of capturing the full experimental lineage of every data point, independent replication becomes computationally and practically intractable.

Without mechanisms to address these dimensions, autonomous science risks becoming non-auditable — generating vast quantities of data and conclusions that cannot be independently evaluated, attributed to specific experimental events, or contested through conventional scientific channels. This is not merely a technical concern but an epistemological one: non-auditable science cannot form the foundation of reliable knowledge, regulatory decisions, or policy-relevant conclusions. The Digital Trust Score (DTS) framework, introduced in Section 3, provides a quantitative model for characterizing and measuring these trust dimensions.

1.3 Scientific Sovereignty Challenges

The global AI research infrastructure is concentrated within a small number of commercial organizations operating primarily in a handful of jurisdictions. The large language models, robotic control algorithms, and data processing platforms that constitute the operational backbone of autonomous research are increasingly developed, owned, and governed by private entities subject to the laws and interests of specific national

contexts. As autonomous research systems become operationally dependent on these platforms, a structural sovereignty problem emerges that has no precedent in the history of scientific organization.

Scientific infrastructure is not merely a technical asset — it is a strategic resource that determines which questions are asked, which discoveries are made, and which actors benefit from the resulting knowledge. When the infrastructure for autonomous discovery is controlled by external commercial entities, several critical risks materialize:

- **Opaque discovery pipelines** — Research conclusions generated by proprietary autonomous systems cannot be fully audited by the institutions commissioning the research, undermining the scientific principle that methodology must be transparent and reproducible.
- **External control of research infrastructure** — Proprietary systems can be withdrawn, modified, or subject to export controls, leaving dependent institutions without access to their own research capabilities with minimal warning.
- **Limited national research autonomy** — Nations that depend on foreign commercial AI platforms for their autonomous research infrastructure cannot guarantee the independence of their scientific conclusions, creating potential conflicts of interest in regulatory, clinical, and policy applications.
- **Data exfiltration risks** — Experimental data generated within national research systems may be transmitted to and processed by foreign cloud infrastructure, raising concerns about intellectual property security and research competitiveness.

These risks are not hypothetical — they are active structural conditions of the current autonomous research landscape. The FRYX Interoperability Protocol, described in Section 4, is designed to provide a sovereignty-preserving alternative: an open, non-proprietary coordination architecture that enables institutions and nations to maintain governance over their autonomous research ecosystems while still benefiting from global collaborative discovery networks.

1.4 Research Objective

This work proposes the Digital Trust & Interoperability Framework (DTIF), a unified architecture designed to address the three structural challenges identified above through a coherent set of integrated components. Specifically, DTIF is designed to:

- Standardize autonomous laboratory infrastructure classification through the SDL Maturity Model, providing a universal vocabulary for describing laboratory autonomy levels across institutional and national contexts
- Establish quantitative trust metrics for AI agents through the Digital Trust Score (DTS) framework, enabling threshold-based governance of autonomous experiment authorization
- Enable interoperable discovery networks through the FRYX Interoperability Protocol, providing a layered, open-standard coordination architecture for heterogeneous autonomous research ecosystems
- Articulate the Discovery Economy model, describing the economic and scientific transformations enabled by widespread DTIF adoption
- Identify the governance implications of autonomous discovery at scale, informing policy development for the responsible operation of sovereign autonomous research ecosystems

The remainder of this paper is organized as follows: Section 2 presents the SDL Maturity Model; Section 3 develops the DTS framework; Section 4 describes the FRYX Interoperability Protocol; Section 5 outlines the autonomous discovery workflow; Section 6 articulates the Discovery Economy model; Section 7 discusses governance implications; and Section 8 presents conclusions. Appendices A through D provide operational specifications for each DTIF component.

2. The Self-Driving Laboratory (SDL) Maturity Model

The concept of the self-driving laboratory — a research system capable of autonomously designing, executing, and interpreting experiments — has emerged as a central paradigm in discussions of the future of scientific discovery. However, the term 'self-driving laboratory' is applied inconsistently across the literature, encompassing systems ranging from simple robotic dispensers executing fixed protocols to fully AI-directed discovery platforms capable of formulating novel research objectives. This terminological ambiguity impedes inter-institutional communication, constrains benchmarking efforts, and complicates governance and regulatory frameworks that must specify which systems require oversight.

The SDL Maturity Model addresses this gap by providing a formal, six-level classification framework for laboratory autonomy. The model draws structural inspiration from analogous maturity taxonomies in adjacent fields — including the SAE International levels of driving automation for autonomous vehicles and the capability maturity model integration (CMMI) framework for software development process maturity — while developing domain-specific criteria appropriate to the unique characteristics of autonomous scientific research.

Each SDL level is defined by the allocation of decision-making authority between human researchers and AI/robotic systems across three axes: the human role in experimental operations, the AI/robotic role in experimental operations, and the nature of the decision loop governing experimental progression. This three-axis definition provides the precision necessary for unambiguous classification of real-world laboratory systems while remaining sufficiently general to accommodate the diversity of experimental domains and disciplinary contexts in which autonomous laboratories operate.

Table 1. SDL Autonomy Classification Framework

Level	Designation	Human Role	AI/Robotic Role	Decision Loop
L0	**Manual**	Full execution	None	Open
L1	**Automated**	Script supervision	Repetitive tasks	Pre-defined
L2	**Assisted**	Approval gatekeeper	Generates options	Human-in-the-loop
L3	**Closed-loop**	Supervisory	Self-correcting cycles	AI-in-the-loop
L4	**Collaborative**	Strategic goals	Multi-agent coordination	Agent-to-agent
L5	**Sovereign**	Policy and ethics only	Self-directed discovery	Autonomous ecosystem

Table 1: The SDL Maturity Model classifies laboratory autonomy across six levels from full human execution (L0) to sovereign autonomous discovery ecosystems (L5). Classification is determined by human role, AI/robotic role, and decision loop architecture.

2.1 Level 0 — Manual Research Systems

Level 0 represents the baseline of conventional laboratory practice, in which all experimental steps are conceived, designed, executed, and evaluated by human researchers without computational automation of decision-making. Laboratory instruments may be electronically controlled and data may be digitally recorded,

but the sequencing of experimental actions and the interpretation of results remain entirely within the domain of human cognition.

L0 systems are characterized by unlimited flexibility — a skilled human researcher can adapt experimental protocols in real-time in response to unexpected observations, apply intuitive reasoning to anomalous data, and draw on contextual knowledge accumulated through years of disciplinary training. These capabilities constitute the enduring comparative advantage of human researchers over automated systems. However, L0 systems are inherently constrained by human cognitive bandwidth, physical endurance, and the variability introduced by inter-operator differences in technique. Throughput is limited to what a small team of researchers can physically execute, typically on the order of dozens of experimental conditions per week.

The L0 designation serves as the reference baseline for the SDL classification system and remains appropriate for highly specialized research domains in which experimental conditions are too variable or novel to be systematically codified for automation.

2.2 Level 1 — Automated Execution

Level 1 systems introduce robotic automation for the physical execution of predefined experimental tasks, while human researchers retain complete authority over protocol design and experimental strategy. Automation at L1 is characterized by rigid scripting — the robotic system executes a fixed sequence of operations specified in advance by human operators and cannot adapt its behaviour in response to experimental outcomes.

Representative L1 implementations include liquid-handling robots programmed to perform fixed dispensing sequences, automated plate readers executing scheduled measurement protocols, and conveyor-belt sample processing systems operating under fixed timing parameters. These systems provide significant throughput advantages over L0 for well-characterized, repetitive experimental workflows while maintaining complete human oversight of experimental strategy.

The primary limitation of L1 systems is their brittleness — unexpected events such as instrument failures, reagent contamination, or sample anomalies cannot be detected or responded to without human intervention. L1 systems cannot modify their behaviour based on intermediate results, and thus cannot participate in closed-loop discovery cycles. Their value is primarily as throughput amplifiers for experimental workflows that have already been validated and standardized.

2.3 Level 2 — AI-Assisted Research

Level 2 systems incorporate AI models capable of analysing experimental data and proposing experimental candidates or protocol optimizations, while human researchers retain final approval authority over all experimental decisions. The human-in-the-loop architecture of L2 systems positions AI as an analytical advisor whose recommendations must be explicitly endorsed before execution.

AI assistance at L2 typically takes the form of predictive models that generate ranked candidate lists for experimental validation — molecular property prediction models that score candidate compounds by predicted bioactivity, Bayesian optimization algorithms that recommend experimental parameter settings likely to optimize a defined objective function, or natural language processing systems that extract experimental insights from scientific literature to guide hypothesis formation.

L2 systems represent a qualitative shift from L1 by introducing data-driven experimental planning. However, the human approval requirement imposes a rate-limiting step on experimental throughput that constrains the

degree to which AI-driven recommendations can be fully exploited. At L2, the computational–experimental paradox is ameliorated but not resolved: AI throughput advantages are partially captured while the epistemic and safety benefits of human oversight are preserved.

2.4 Level 3 — Closed-Loop Discovery

Level 3 represents a critical transition point in the SDL hierarchy: the introduction of AI-in-the-loop experimental control, in which the system autonomously designs and executes experiments in response to incoming data without requiring human approval for individual experimental decisions. At L3, the research system forms a self-correcting discovery cycle in which experimental results continuously update the AI model's understanding of the experimental space, which in turn drives subsequent experimental design decisions.

The closed-loop architecture of L3 enables continuous operation without the rate-limiting step imposed by human approval cycles. Experimental campaigns can proceed around the clock, with the AI system operating across time zones and working patterns that no human research team could match. This capability allows L3 systems to explore experimental parameter spaces with a thoroughness and efficiency that creates qualitatively new discovery opportunities.

Human oversight at L3 transitions from operational management to supervisory monitoring: researchers define research objectives, establish safety constraints, and review system-generated progress reports, but do not approve individual experimental cycles. Intervention occurs when predefined anomaly thresholds are exceeded — instrument failures, unexpected result patterns, or safety-relevant observations trigger human review and system pause. The governance of L3 systems requires robust anomaly detection mechanisms and clear escalation protocols, both of which are formalized in the DTS framework and FRYX Governance Layer.

2.5 Level 4 — Collaborative Multi-Agent Research

Level 4 architectures extend the closed-loop discovery paradigm to networks of multiple coordinated autonomous agents, each contributing specialized capabilities to a distributed discovery workflow. At L4, the research system is no longer a single laboratory operating autonomously but a coordinated ensemble of autonomous laboratories, computational agents, and analytical systems that collectively address research objectives of greater scope and complexity than any single agent could achieve independently.

Multi-agent coordination at L4 requires communication protocols that enable agents to share experimental findings, negotiate task assignments, resolve conflicting evidence, and collectively update a shared model of the experimental space. The FRYX Interoperability Protocol provides the architectural foundation for L4 operation, defining the communication standards, data formats, and coordination mechanisms necessary for heterogeneous agents to function as a coherent research ensemble.

Human researchers at L4 provide high-level strategic direction — defining the research domain, establishing ethical boundaries, and setting performance objectives — while the agent network autonomously manages the operational details of experimental execution. The governance requirements of L4 systems are substantially more complex than those of L3, as conflicts between agent recommendations, resource allocation decisions, and knowledge integration protocols must be managed without direct human involvement in routine operations.

2.6 Level 5 — Sovereign Autonomous Discovery

Level 5 represents the theoretical and aspirational apex of the SDL classification: fully autonomous scientific ecosystems capable of self-directed discovery across the complete research lifecycle, from the formulation of research objectives through experimental execution, result interpretation, knowledge dissemination, and the identification of new research directions. At L5, the system is not merely executing a research programme defined by human researchers but is actively participating in the definition of the research agenda itself.

Human involvement in L5 ecosystems is limited to the articulation of ethical constraints and overarching policy boundaries — the system must not pursue research objectives that conflict with defined ethical principles, national security requirements, or regulatory frameworks. Within these boundaries, the L5 system autonomously determines which scientific questions are most productive to investigate, allocates research resources across experimental domains, and manages the full complexity of a self-sustaining scientific enterprise.

L5 systems do not currently exist in fully realized form, but the trajectory of autonomous research development — from L1 automation to increasingly sophisticated L3 and nascent L4 systems — suggests that L5 capability represents a plausible medium-term technological objective rather than a purely speculative future state. The governance frameworks, trust architectures, and interoperability standards established by DTIF are intended to provide the infrastructure necessary to enable responsible L5 development when the underlying technical capabilities mature.

It is important to note that the SDL Maturity Model is descriptive, not prescriptive: higher levels are not intrinsically superior to lower levels for all research contexts. The appropriate SDL level for a given laboratory depends on the nature of the research domain, the degree to which experimental protocols can be systematically formalized, the availability of validated AI models, and the institutional capacity for autonomous system governance. The model provides a shared vocabulary for characterizing capability, not a mandate for capability escalation.

3. Digital Trust Scoring (DTS) Framework

The deployment of autonomous research agents in operational scientific settings requires a principled basis for evaluating the reliability of their outputs. In human research communities, trust in scientific findings is established through a distributed social process — peer review, independent replication, and the accumulation of consistent evidence across multiple laboratories and research groups. This social process of trust accumulation is slow, costly, and imperfect, but it provides a broadly accepted epistemic foundation for scientific knowledge claims.

Autonomous research agents, operating at speeds and scales that preclude conventional peer review timelines, require a complementary trust framework that can provide real-time, quantitative assessments of agent reliability. The Digital Trust Score (DTS) framework provides precisely this: a composite metric derived from four operationally measurable parameters that collectively characterize the dimensions of trustworthiness most critical to autonomous research contexts.

The DTS is designed to be: computationally tractable in real-time operational settings; interpretable by both human overseers and other autonomous agents; sensitive to the specific failure modes most characteristic of AI-driven research systems; and extensible to accommodate new trust dimensions as autonomous research practice evolves.

3.1 DTS Formula and Parameter Overview

The Digital Trust Score is defined by the following composite formula:

$$T = (V \times O \times Tr) / U$$

Equation 1: Digital Trust Score (DTS) — composite reliability metric for autonomous research agents

Table 2. DTS Parameter Definitions and Measurement Specifications

Parameter	Definition	Measurement Proxies	Range
Verifiability (V)	Proportion of steps independently reproducible	Cross-lab replication rate, protocol conformance score	0.0 – 1.0
Observability (O)	Transparency of AI inference pathways	SHAP attribution coverage, attention map availability	0.0 – 1.0
Traceability (Tr)	Integrity of data provenance chain	Cryptographic audit chain completeness	0.0 – 1.0
Model Uncertainty (U)	Hallucination and instability risk	Ensemble disagreement index, calibration error rate	0.0 – ∞ (denominator)

Table 2: The four DTS parameters, their operational definitions, recommended measurement proxies, and valid ranges. The numerator parameters (V, O, Tr) reward reliability while the denominator (U) penalizes uncertainty.

The multiplicative structure of the numerator — rather than an additive formulation — is intentional: a system that scores highly on two of the three numerator parameters but fails catastrophically on the third will receive

a substantially lower composite score than an additive formulation would assign. This reflects the operational reality that autonomous research systems must be reliable across all trust dimensions simultaneously; a highly transparent but non-reproducible system, or a reproducible but opaque one, remains fundamentally untrustworthy for high-stakes scientific applications.

3.2 Verifiability

The Verifiability index (V) measures the proportion of experimental steps executed by an autonomous research agent that can be independently reproduced by external agents or automated verification systems. Verifiability addresses the most fundamental requirement of scientific methodology: that experimental claims must be reproducible by parties other than those who originally generated them.

Verifiability assessment encompasses three operational sub-dimensions. Protocol transparency quantifies the degree to which the autonomous system's experimental protocols are expressed in standardized, machine-readable formats accessible to external validation agents — proprietary or undocumented protocols receive a proportionally reduced transparency score. Instrument calibration integrity tracks the continuous calibration status of all measurement instruments deployed in the experimental workflow, with any uncalibrated instrument state reducing the verifiability score for measurements acquired during that interval. Inter-laboratory replication success rate measures the empirical proportion of independently attempted replications that produce results within the defined tolerance bounds of the original experimental outputs.

For a newly deployed autonomous research agent with no replication history, V is computed from the first two sub-dimensions only, with the replication rate contributing as historical data accumulates. This design enables the DTS framework to assign provisional trust scores to new agents while preserving the expectation of empirical validation through the replication sub-dimension over time.

3.3 Observability

The Observability index (O) measures the degree to which the reasoning processes of an AI agent operating within an autonomous research system are accessible and interpretable to external evaluators. Observability is the DTS framework's primary response to the opacity problem characteristic of deep learning systems deployed in scientific contexts.

Observability is assessed across three sub-dimensions. Inference pathway exposure quantifies the proportion of the agent's decision chain — from input data to experimental recommendation — that is expressed in human-interpretable or machine-auditable form. Feature attribution completeness measures the proportion of input features that can be attributed a quantitative contribution to the agent's output using established interpretability methods such as SHAP (SHapley Additive exPlanations) values or integrated gradient attribution. Confidence distribution accessibility evaluates whether the agent exposes calibrated probability distributions over its outputs, enabling downstream evaluators to distinguish high-confidence recommendations from exploratory suggestions.

Systems built on interpretable model architectures — Bayesian networks, Gaussian process regressors, decision trees, or attention-mechanism-enhanced transformers with accessible attention maps — receive higher baseline Observability scores than black-box ensembles or uninterpreted deep networks. The DTS framework thus creates a positive governance incentive for the development and deployment of interpretable autonomous research architectures.

3.4 Traceability

The Traceability constant (Tr) measures the completeness and integrity of the data provenance chain maintained by an autonomous research agent, from raw instrument sensor output through analytical processing steps, AI inference stages, and final knowledge integration. Traceability addresses the requirement that every data point in a scientific record must be attributable to a specific, time-stamped experimental event, with all transformations applied to that data systematically documented.

Traceability is operationalized through three sub-dimensions. Provenance chain completeness quantifies the proportion of final research outputs for which a complete, unbroken provenance chain can be constructed linking each data point to its originating experimental event. Audit log cryptographic integrity measures whether the provenance chain is secured against post-hoc modification through cryptographic mechanisms — hash-linked audit chains, immutable distributed ledger records, or digitally signed experimental records. Transformation documentation coverage assesses the proportion of data processing transformations — normalization, outlier filtering, dimensionality reduction — that are documented in a standardized, machine-readable format that enables independent recomputation.

High Traceability scores are particularly critical for research outputs intended for regulatory submission — pharmaceutical clinical trial data, materials certification testing, and environmental monitoring results — where demonstrable provenance integrity is a legal requirement as well as a scientific one.

3.5 Model Uncertainty

Model uncertainty (U) quantifies the composite risk of hallucination, probabilistic instability, and out-of-distribution inference failure within an autonomous research agent's AI component. As the denominator of the DTS formula, elevated uncertainty suppresses the composite Trust Score, ensuring that agents operating in high-uncertainty regimes cannot achieve high trust ratings regardless of their performance on the numerator dimensions.

Model uncertainty is assessed through three complementary metrics. Ensemble disagreement index measures the variance in outputs produced by an ensemble of independently trained versions of the same model architecture when applied to identical inputs — high variance indicates that the model is operating in a regime where its training data provides insufficient constraint on output predictions. Calibration error quantifies the discrepancy between the model's expressed confidence levels and its empirical accuracy across a validation dataset — a model that reports 90% confidence on predictions it is correct on only 60% of the time has high calibration error and therefore high uncertainty. Out-of-distribution detection rate measures the proportion of input cases that the model correctly identifies as falling outside its reliable operating range, triggering human escalation rather than autonomous experimental action.

3.6 Trust Thresholds

The DTS framework establishes a tiered threshold system that governs the authorization of autonomous experiments based on their risk profile and the computed DTS of the executing agent. Trust thresholds are defined at the institutional and regulatory level, with higher-risk experimental domains requiring elevated DTS values before autonomous execution authorization is granted.

Table 3. DTS Trust Threshold Classification for Autonomous Research Authorization

Research Class	Description	Min DTS	Governance Requirement
Exploratory Research	Low-stakes hypothesis screening	**$T \geq 0.40$**	Informational only
Confirmatory Research	Validation of prior findings	**$T \geq 0.60$**	Requires human review
Regulated Research	Clinical, radiological, chemical synthesis	**$T \geq 0.80$**	Dual-agent verification
Sovereign Discovery	Policy-critical knowledge generation	**$T \geq 0.95$**	Governance board approval

Table 3: Trust threshold requirements for autonomous experiment authorization across research risk classes. Higher-risk domains require elevated DTS values and additional governance controls before autonomous execution is permitted.

Threshold enforcement is implemented through the Digital Trust Engine component of the FRYX Governance and Trust Layer, described in Section 4.3. The threshold system is designed to be dynamically adjustable by institutional governance bodies in response to evolving regulatory requirements, empirical evidence of agent performance, or changes in the risk profile of experimental domains. Threshold parameters are encoded as policy-as-code specifications that can be updated without system downtime, enabling responsive governance of dynamic autonomous research environments.

4. FRYX Interoperability Protocol

The FRYX Interoperability Protocol provides the communications and coordination architecture that enables autonomous laboratories operating under DTIF to function as a coherent distributed research network. The name FRYX derives from the conceptual framework of a flexible research exchange — an open, non-proprietary coordination protocol analogous in function to TCP/IP for internet communication or HTTP for web-based data exchange, but optimized for the specific data types, latency requirements, and governance constraints characteristic of autonomous scientific research.

The FRYX architecture addresses a critical gap in the current autonomous research landscape: the absence of a universal protocol enabling autonomous laboratories built on different hardware platforms, controlled by different AI systems, and governed by different institutional policies to coordinate their research activities and share validated findings. In the absence of such a protocol, each autonomous laboratory operates as an isolated silo — generating data and conclusions that are inaccessible to other laboratories without bespoke integration effort. FRYX transforms these isolated silos into a global network of interoperable research nodes, enabling collective intelligence at a scale that no individual laboratory can achieve.

Table 4. FRYX Protocol Layer Architecture Summary

Layer	Function	Key Components	Integration Points
Physical Execution	Hardware abstraction & actuation	Instrument API gateways, robotic controllers, sensor capture	Hardware manufacturers, IoT protocol stacks
Logic & Intelligence	AI coordination & planning	Hypothesis engine, model translator, multi-agent coordinator	LLM providers, scientific AI frameworks
Governance & Trust	Compliance & integrity	Digital Trust Engine, policy-as-code, cross-agent validation	Regulatory bodies, institutional ethics boards

Table 4: The three-layer FRYX architecture assigns distinct functional responsibilities to hardware abstraction, AI coordination, and governance enforcement, enabling modular implementation and layer-independent upgrades.

4.1 Physical Execution Layer

The Physical Execution Layer constitutes the lowest tier of the FRYX architecture, providing standardized interfaces between the FRYX protocol stack and the physical laboratory hardware through which experiments are executed. The fundamental design principle of this layer is hardware agnosticism: the layer must be implementable across the full diversity of laboratory instrument manufacturers, robotic platform designs, and sensor modalities present in operational autonomous research environments without requiring proprietary modifications to existing equipment.

4.1.1 Instrument API Gateways

Instrument API gateways are the primary integration mechanism through which FRYX communicates with laboratory hardware. Each gateway translates the FRYX standardized instrument command schema — a

universal data format specifying instrument actions, measurement parameters, timing requirements, and data return specifications — into the proprietary communication protocol of a specific instrument or instrument class. Gateway implementations are maintained as open-source software modules within the FRYX instrument gateway registry, enabling institutions to deploy interoperability with a growing library of supported instruments without requiring bespoke integration development for each deployment.

4.1.2 Robotic Experiment Controllers

Robotic experiment controllers are embedded systems that govern the physical movements and operational states of robotic laboratory platforms — articulated arm robots, cartesian liquid-handling systems, mobile laboratory robots, and specialized automated synthesis platforms. The FRYX robotic controller interface specifies a standardized motion and actuation command language, enabling the FRYX Logic and Intelligence Layer to specify experimental operations in terms of abstract laboratory actions (aspirate 50 µL from well A1; incubate at 37°C for 24 hours) rather than platform-specific motor control sequences. Controller firmware translates these abstract specifications into platform-specific execution, isolating the intelligence layer from hardware implementation details.

4.1.3 Sensor Data Capture Systems

Real-time sensor data capture systems aggregate measurement outputs from all instruments and environmental monitoring systems within the autonomous laboratory into structured experimental records compliant with the FRYX data schema. The FRYX sensor data capture specification defines a universal metadata standard that accompanies all raw measurement data: instrument identifier, calibration status, operator (human or AI) authorization token, timestamp with cryptographic attestation, environmental conditions, and a hash linking the record to the FRYX audit chain. This metadata standard is the foundation of the Traceability dimension of the DTS framework.

4.2 Logic and Intelligence Layer

The Logic and Intelligence Layer constitute the cognitive centre of FRYX-compliant autonomous research systems, coordinating the AI models and experimental planning functions that drive discovery. This layer is responsible for translating scientific objectives defined at the strategic level into concrete experimental programmes, managing the flow of data between AI systems and experimental execution platforms, and coordinating the activities of multiple AI agents operating within a distributed discovery network.

4.2.1 Hypothesis Generation Engine

The hypothesis generation engine integrates data from multiple sources — validated experimental records, scientific literature embeddings, domain knowledge graphs, and real-time experimental results — to produce ranked lists of candidate hypotheses for experimental validation. The engine supports multiple model architectures, enabling institutions to deploy hypothesis generation capabilities appropriate to their research domain: large language models for literature-grounded hypothesis synthesis, Bayesian optimization algorithms for parameter space exploration, graph neural networks for molecular and materials property prediction, and reinforcement learning agents for protocol optimization. The FRYX universal model interface enables these heterogeneous architectures to contribute to a unified hypothesis generation pipeline without requiring a common underlying model format.

4.2.2 Universal Model Translator

The universal model translator is a semantic interoperability module enabling AI models with heterogeneous input/output specifications to exchange scientific data without requiring bespoke integration. The translator operates on a shared scientific data ontology — a formal specification of the entities, relationships, and properties relevant to autonomous research across experimental domains — mapping proprietary model input/output schemas to ontology-compliant representations that any FRYX-connected model can consume. This eliminates the integration barrier that currently prevents scientific AI systems from different vendors or research groups from collaborating within a unified discovery workflow.

4.2.3 Multi-Agent Coordination System

The multi-agent coordination system manages the distributed task allocation and consensus mechanisms required for L4-class collaborative research at network scale. The coordination system implements a contract-net protocol for task allocation — autonomous agents broadcast experimental task requirements, receive bid proposals from capable agents, and select execution agents based on computed DTS scores, declared capacity, and geographic or regulatory constraints. Conflict resolution between agents that produce incompatible experimental conclusions is managed through a structured evidence-weighting protocol that integrates DTS scores, provenance quality assessments, and inter-agent consensus metrics to determine which conclusions are incorporated into the shared knowledge base.

4.3 Governance and Trust Layer

The Governance and Trust Layer constitute the uppermost tier of the FRYX architecture, providing the enforcement mechanisms through which institutional and regulatory policies govern autonomous research operations. This layer is the operational expression of the DTS framework, translating trust scores and threshold specifications into real-time access controls, experiment authorization decisions, and compliance audit records.

4.3.1 Digital Trust Engine

The Digital Trust Engine is a real-time computation and monitoring system that continuously evaluates the DTS of all agents operating within a FRYX-connected research network. The engine maintains a persistent trust state for each registered agent, updating DTS scores as new calibration data, replication records, and observability assessments become available. When an agent requests authorization to execute an experiment, the Trust Engine computes the current DTS, compares it against the threshold requirements defined for the experiment's risk class, and either grants authorization or initiates a human escalation process. All Trust Engine decisions are logged to the FRYX audit chain with cryptographic attestation, providing a complete, tamper-evident record of governance decisions.

4.3.2 Policy-as-Code Enforcement

Policy-as-code enforcement translates institutional and regulatory research governance policies into machine-executable rules applied at the point of experiment authorization. FRYX policy specifications are expressed in a formal policy language designed to accommodate the full diversity of national and institutional research governance frameworks — biosafety regulations, controlled substance research restrictions, ethics board decisions, intellectual property embargoes, and export control requirements. Policy specifications are versioned, cryptographically signed by authorized governance authorities, and automatically applied by the Trust Engine without manual intervention, ensuring consistent compliance across all autonomous operations within the governed network.

4.3.3 Cross-Agent Validation Protocols

Cross-agent validation protocols are distributed verification mechanisms enabling independent autonomous agents to validate each other's experimental claims before those claims are integrated into the shared scientific knowledge base. Validation requests specify the claim to be validated, the evidence base supporting it, and the experimental protocol required for independent verification. Validation agents independently execute the verification protocol and submit their results to a consensus evaluation process that weights individual validation outcomes by the validating agent's DTS. Claims that achieve a configurable consensus threshold across multiple independent validating agents are certified as validated findings and integrated into the shared knowledge base; claims that fail consensus verification are flagged for human review.

5. Autonomous Discovery Workflow

The integration of the SDL Maturity Model, DTS framework, and FRYX Interoperability Protocol into a coherent operational system is realized through the Autonomous Discovery Workflow — a structured cycle that governs the progression of scientific inquiry from initial hypothesis formulation through knowledge integration and iterative refinement. The workflow is designed to be executable at all SDL levels from L2 upward, with the degree of human involvement scaling inversely with SDL level.

The canonical DTIF discovery cycle consists of six sequential stages, each corresponding to a defined set of interactions between the FRYX protocol layers and the institutional governance framework. The cyclical nature of the workflow is its most significant architectural feature: each completed cycle generates outputs that improve the quality of subsequent cycles, creating a compounding discovery dynamic that becomes more efficient as the system accumulates validated experimental history.

1. **AI Hypothesis Generation —** The Logic and Intelligence Layer synthesize available experimental data, validated literature embeddings, and domain knowledge graph information to produce a ranked list of candidate hypotheses for experimental validation. The hypothesis generation engine applies the current state of the research agent's predictive models to identify experimental candidates most likely to advance the defined research objective, weighted by predicted information gain and estimated experimental cost. The output of this stage is a structured hypothesis manifest: a machine-readable specification of each candidate experiment, its predicted outcome distribution, and its estimated contribution to the research objective.
2. **Multi-Agent Evaluation** — The hypothesis manifest is submitted to the multi-agent coordination system for network-wide evaluation. Connected agents assess the hypothesis manifest against their own knowledge states, experimental capabilities, and current resource availability, producing bid proposals that specify their capacity to execute elements of the experimental programme. The coordination system allocates experimental tasks across available agents based on DTS scores, declared specialization, and workload balancing considerations. At L4 SDL level, this stage may involve dozens of geographically distributed laboratories negotiating task assignments across institutional and national boundaries in real-time.
3. **Robotic Experiment Execution** — Authorized experimental protocols are transmitted to the Physical Execution Layer of the assigned laboratories, which actuates laboratory instruments and robotic platforms to execute the specified procedures. The Digital Trust Engine verifies that the executing agent holds a current DTS meeting the threshold requirements for the assigned experiment class before authorizing Physical Execution Layer actuation. All physical execution events are logged to the FRYX audit chain with cryptographic attestation of instrument states, operator identities, and environmental conditions at the time of execution.
4. **Sensor Data Acquisition** — Real-time sensor systems capture experimental outputs across all relevant measurement modalities — spectroscopic, chromatographic, imaging, electrochemical, genomic, and environmental — and transmit structured data records conforming to the FRYX sensor data schema to the Logic and Intelligence Layer. Data acquisition is accompanied by automated quality

control assessments that flag records deviating from instrument-specific quality thresholds, triggering human review before such records are incorporated into the experimental dataset.

5. **Trust Verification** — The Digital Trust Engine computes updated DTS values for all agents that contributed to the experimental cycle, incorporating new calibration, replication, and observability assessments generated during execution. Experimental data records are subjected to the cross-agent validation protocol: independent validation agents attempt replication or computational verification of key findings, and the consensus validation status of each finding is recorded. Findings that achieve validated consensus status at the appropriate threshold for their risk class are certified for knowledge integration; findings that do not achieve consensus are quarantined for human review.
6. **Knowledge Integration** — Validated experimental findings are integrated into the shared scientific knowledge base, updating the hypothesis generation engine's predictive models with new training data, enriching the domain knowledge graph with validated relationships, and revising the ranked hypothesis manifest for the next discovery cycle. The knowledge integration stage also updates the institutional intellectual property registry with new validated discoveries, initiating patent or publication preparation workflows as specified by institutional policy-as-code specifications. The completion of the knowledge integration stage initiates the next discovery cycle, with the updated knowledge state informing a higher-quality hypothesis generation process.

The six-stage DTIF discovery cycle represents a significant departure from conventional research workflows in several respects. The automated governance mechanisms embedded in stages 3, 4, and 5 eliminate the administrative overhead that currently constitutes a substantial fraction of research cycle time — regulatory compliance checks, data quality reviews, and cross-laboratory coordination are handled programmatically rather than through manual institutional processes. The compounding improvement dynamic of the cyclical architecture enables the system to become progressively more efficient over time, with each validated experimental cycle reducing the uncertainty in predictive models and thereby improving the quality and efficiency of subsequent hypothesis generation. These properties collectively enable the DTIF discovery cycle to realize the throughput advantages of autonomous research at scale while maintaining the governance standards necessary for trustworthy scientific practice.

6. The Discovery Economy Model

The widespread deployment of DTIF-compliant autonomous laboratories creates the conditions for a fundamentally new scientific and economic paradigm: the Discovery Economy. In this model, distributed autonomous research infrastructures function not merely as tools for generating academic knowledge but as continuous generators of intellectual property, commercially actionable data assets, and validated scientific discoveries with direct economic value. The Discovery Economy represents a structural transformation in the relationship between scientific research and economic value creation — one with profound implications for national research policy, institutional funding models, and the global distribution of scientific capacity.

6.1 From Cost Centers to Discovery Infrastructure

Under conventional research economics, laboratory infrastructure is classified as an operational cost center — a resource-consuming enterprise that generates knowledge as a non-tradable public good, with economic returns realized indirectly through publication prestige, technology licensing, and talent development over multi-decade timescales. This model has produced extraordinary scientific achievements but has increasingly strained institutional funding models as the cost of state-of-the-art research equipment, skilled personnel, and regulatory compliance has escalated faster than public research budgets.

DTIF-compliant autonomous laboratories fundamentally transform this economic model. By operating continuously without labour costs proportional to throughput, and by generating validated experimental findings at scales that create commercially licensable IP portfolios as a routine operational output, autonomous laboratories can transition from cost centers to revenue-generating discovery infrastructure. Pharmaceutical companies have already demonstrated the precursor to this model through their investment in high-throughput screening facilities — the DTIF extends and generalizes this economic logic to the full diversity of research domains accessible to autonomous experimentation.

The key enabling conditions for this transition are: (1) the FRYX interoperability architecture, which enables experimental output to be formatted as tradeable data assets that retain full provenance certification; (2) the DTS framework, which provides the trust certification necessary for commercial and regulatory acceptance of autonomously generated research findings; and (3) the SDL Maturity Model, which enables institutional investors and licensing partners to accurately assess the autonomy capability and therefore the production capacity of a given research system.

6.2 Regional Autonomous Research Hubs

One of the most significant economic implications of DTIF adoption is the potential to democratize high-value scientific discovery by reducing the capital and organizational requirements for operating state-of-the-art research infrastructure. The FRYX Protocol's hardware-agnostic architecture enables small-scale autonomous laboratories — operating within regional universities, national research institutes, or specialized private research organizations — to generate high-value discoveries without requiring the massive capital investments that have historically concentrated scientific capacity within a small number of elite institutions in wealthy nations.

Regional autonomous research hubs can adopt domain-specialization strategies that leverage local scientific expertise, natural resources, or industrial context to create competitive advantages in specific discovery sectors. A regional hub with expertise in tropical plant biochemistry can operate an autonomous natural product

discovery programme that generates globally valuable pharmaceutical lead compounds without requiring investment in the full breadth of research infrastructure associated with a major research university. The DTIF's trust certification and interoperability architecture enables the hub's discoveries to be integrated into global pharmaceutical development pipelines with full provenance transparency, creating revenue streams that support sustainable research operations.

The emergence of regional autonomous research hubs represents a structural decentralization of scientific capacity with significant implications for research equity. Nations that have historically been research consumers — importing scientific knowledge generated in wealthy research centres — can transition to research producers by investing in DTIF-compliant autonomous research infrastructure rather than attempting to replicate the full institutional model of established research universities. This transition represents one of the most significant opportunities for the advancement of global scientific equity in the modern era.

6.3 Global Networks of Autonomous Research Nodes

At scale, DTIF-compliant laboratories aggregate into global networks of interoperable autonomous research nodes — distributed intelligence systems capable of coordinating discovery activities across institutional and national boundaries with a coherence and efficiency unattainable by conventional international research collaboration. The FRYX Interoperability Protocol provides the technical foundation for these networks, enabling nodes to share experimental findings, coordinate hypothesis exploration, and validate each other's conclusions within a shared trust framework.

Global autonomous research networks create discovery economies of scale that do not exist in conventional international research collaboration. In conventional collaboration, the costs of coordinating experimental activities across laboratories — protocol standardization, data format reconciliation, inter-laboratory communication, and credit attribution — often approach or exceed the scientific value of the collaboration itself. FRYX automation of these coordination functions enables distributed networks to function with a coordination overhead that scales sub-linearly with network size, creating accelerating returns to scale as the network grows.

Table 5. Discovery Economy Sector Applications and Economic Impact Projections

Sector	Application Domain	Autonomous Capability	Economic Impact
Pharmaceutical	Drug discovery, toxicology screening	High-throughput molecular validation	Accelerated IND filing, reduced clinical failure rates
Materials Science	Novel alloy, polymer, semiconductor discovery	Property-to-structure optimization	Reduced materials development cycles from years to weeks
Agricultural Biotech	Crop yield, disease resistance genetics	Phenotypic screening at scale	Food security resilience, reduced chemical inputs
Energy Research	Catalyst discovery, battery materials	Electrochemical cycle optimization	Faster path to grid-scale storage and clean energy
Genomics	Gene-function mapping, CRISPR optimization	Sequence-activity relationship modelling	Precision medicine personalization at population scale

Table 5: Illustrative sector applications of DTIF-enabled autonomous discovery, with representative autonomous capabilities and projected economic impact characteristics across five high-value research domains.

The governance of global autonomous research networks requires international frameworks that do not currently exist in fully realized form. The FRYX Governance and Trust Layer provides the technical architecture for sovereignty-preserving cross-border research collaboration — enabling nodes in different national jurisdictions to participate in shared discovery workflows while maintaining compliance with their respective regulatory environments and protecting their intellectual property interests. However, the realization of effective global network governance will ultimately require international treaty frameworks, mutual recognition agreements for DTS certification standards, and shared protocols for the attribution and commercial exploitation of network-generated discoveries.

7. Implications for Scientific Governance

The deployment of autonomous discovery systems at the scale envisioned by the Discovery Economy model necessitates commensurate advances in scientific governance infrastructure. Governance frameworks developed for human-operated laboratories — grounded in the accountability of identifiable researchers, the legibility of manual experimental records, and the relatively slow pace of experimental progress — are not adequate for the operational realities of L3–L5 autonomous research systems. New governance architectures are required that can operate at the speed of autonomous research while maintaining the accountability, reproducibility, and ethical compliance standards that legitimate scientific practice demands.

The DTIF identifies three foundational governance requirements for the responsible operation of sovereign autonomous research ecosystems. These requirements are not merely recommendations — they are structural prerequisites for the Discovery Economy model to generate scientific knowledge that is trustworthy, equitable, and epistemically legitimate.

7.1 Open Interoperability Standards

The first governance requirement is the establishment and maintenance of open, non-proprietary interoperability standards governing data formats, communication protocols, trust metrics, and governance specifications within autonomous research networks. Open standards serve multiple governance functions simultaneously: they prevent the monopolization of autonomous research infrastructure by any single commercial or governmental actor; they enable meaningful external audit of autonomous research systems by parties independent of the system operator; they create competitive markets for autonomous research tools and services; and they enable regulatory bodies to develop meaningful oversight frameworks for a technology landscape that would otherwise fragment into incompatible proprietary ecosystems.

The FRYX Interoperability Protocol is designed as an open standard, with all specification documents, data schemas, and reference implementations published under open licensing frameworks that permit free adoption, modification, and redistribution without commercial restriction. The governance of the FRYX standard itself requires an independent standards body — analogous to the Internet Engineering Task Force (IETF) for internet protocols or the World Wide Web Consortium (W3C) for web standards — that can manage the evolution of the specification while maintaining backward compatibility and ensuring that commercial interests do not capture the standards process.

The governance of open standards for autonomous research also requires active enforcement mechanisms. Without explicit policy requirements mandating the use of open standards in publicly funded research infrastructure, institutional procurement decisions may favour proprietary systems with greater short-term capabilities, progressively locking research infrastructure into commercial ecosystems that undermine the long-term governance objectives of openness and sovereignty. Public funding bodies, national research councils, and international research funding organizations have a critical role to play in establishing and enforcing open standards requirements as conditions of research infrastructure funding.

7.2 Transparent AI Governance

The second governance requirement is mandatory transparency in the AI systems deployed within autonomous research infrastructures. Transparent AI governance encompasses multiple dimensions: disclosure of model architectures and training data provenance, publication of model performance benchmarks on standardized

evaluation datasets, mandatory Observability index assessment for all agents operating above DTS threshold L2, and regular third-party audits of deployed autonomous research systems by qualified independent reviewers.

Transparent AI governance is not merely a technical requirement — it is a fundamental condition for maintaining the social legitimacy of autonomous science. When the conclusions of autonomous research systems inform regulatory decisions, clinical practice, or public policy, the populations affected by those decisions have a legitimate claim to understand the basis on which the conclusions were generated. Black-box AI systems that generate scientifically significant conclusions without human-interpretable justification represent a fundamental threat to the epistemological transparency that democratic accountability requires of science.

The DTS Observability index provides a quantitative basis for transparent AI governance requirements, enabling regulatory bodies to specify minimum Observability scores for AI systems deployed in regulated research contexts. Systems that cannot achieve required Observability thresholds are restricted from operating in high-stakes research domains, creating a direct economic incentive for the development of more interpretable autonomous research architectures. This governance mechanism is more effective than prescriptive architectural requirements because it is outcome-focused rather than implementation-focused, accommodating the rapid pace of AI architecture innovation while maintaining the governance objective of interpretability.

7.3 Global Verification Protocols

The third governance requirement is the establishment of globally recognized verification protocols for autonomous research claims — internationally ratified frameworks specifying the standards of independent validation that autonomous research findings must achieve before they can be accepted as certified scientific knowledge by regulatory bodies, publication venues, and policy institutions worldwide.

Global verification protocols must address several technically and politically complex requirements. They must be sufficiently rigorous to provide genuine epistemic assurance of validated findings while remaining practically implementable by research systems operating across diverse experimental domains, institutional contexts, and regulatory environments. They must be internationally recognized by a sufficiently broad coalition of scientific nations to prevent the emergence of fragmented verification regimes that would balkanize the global scientific knowledge base. And they must include mechanisms for the resolution of conflicting validation results — cases in which independent validation attempts by geographically and institutionally diverse agents produce inconsistent conclusions.

The FRYX cross-agent validation protocol provides the technical foundation for global verification, but its translation into internationally recognized standards requires political engagement at the level of international scientific governance bodies — the International Council for Science (ICSU), the International Organization for Standardization (ISO), and national standards bodies in major research nations. The development of these international frameworks is a long-term governance project that will require sustained engagement across scientific, policy, and diplomatic channels. The DTIF positions itself as the technical reference architecture that such international frameworks can adopt, providing a concrete and operational basis for governance discussions that might otherwise remain at an abstract level.

7.4 Ethical and Equity Considerations

Beyond the three structural governance requirements, the deployment of autonomous discovery systems at global scale raises profound ethical and equity considerations that governance frameworks must explicitly address. The potential for autonomous research to dramatically accelerate scientific discovery creates both extraordinary opportunities and serious risks for equity in the global distribution of scientific capacity and benefit.

On the equity dimension, the Discovery Economy model has the potential to either dramatically expand access to scientific discovery capabilities for historically underserved research communities or to further concentrate scientific capacity in well-resourced institutions and nations — depending critically on whether the FRYX open standards architecture is successfully adopted as the dominant coordination framework or is displaced by proprietary alternatives. Governance frameworks that mandate open standards in publicly funded research, provide capacity-building support for regional autonomous research hub development in low- and middle-income countries, and establish benefit-sharing frameworks for discoveries made within global autonomous research networks are prerequisites for realizing the equity potential of the Discovery Economy.

On the ethical dimension, the delegation of scientific decision-making authority to autonomous agents raises questions of accountability that existing ethical frameworks have not been designed to address. When an autonomous research agent makes an experimental decision that produces harmful results — whether through incorrect experimental conclusions, safety protocol failures, or unanticipated applications of validated discoveries — the attribution of responsibility is genuinely novel and requires careful governance design. The DTIF's audit chain architecture provides the evidentiary foundation for accountability investigations, but institutional and regulatory frameworks for assigning and enforcing responsibility in autonomous research contexts must be developed in parallel with the technical architecture.

8. Conclusion

This paper has introduced the Digital Trust & Interoperability Framework (DTIF) as a foundational architecture for the next generation of autonomous scientific research infrastructure. The framework responds to a set of structural challenges — the computational–experimental paradox, the trust gap in autonomous research, and the scientific sovereignty risks of proprietary AI dependence — that collectively constrain the realization of autonomous discovery's transformative potential. Through three integrated components, DTIF provides the architectural foundation necessary to address these challenges comprehensively and coherently.

The Self-Driving Laboratory (SDL) Maturity Model provides a universal vocabulary for characterizing laboratory autonomy across six precisely defined levels, enabling researchers, institutions, funders, and regulators to communicate about autonomous research capabilities with the precision that effective governance requires. By drawing on analogous maturity taxonomy frameworks in adjacent fields and adapting their classification logic to the specific characteristics of autonomous scientific research, the SDL model provides immediate practical utility while establishing a conceptual foundation for long-term standardization efforts.

The Digital Trust Score (DTS) framework establishes a quantitative basis for evaluating the reliability of autonomous research agents through four operationally measurable parameters — Verifiability, Observability, Traceability, and Model Uncertainty — combined in a composite formula that rewards multidimensional reliability and penalizes uncertainty. The tiered trust threshold system derived from DTS scores enables threshold-based governance of autonomous experiment authorization, translating abstract trustworthiness assessments into concrete access control decisions that can be operationally enforced by the FRYX Digital Trust Engine. The DTS framework creates positive governance incentives for the development of interpretable, auditable, and reproducible autonomous research architectures by directly linking trust scores to operational permissions.

The FRYX Interoperability Protocol provides a layered, open-standard coordination architecture for heterogeneous autonomous research ecosystems, enabling the formation of global networks of interoperable research nodes without requiring centralized control or proprietary integration. The three-tier architecture — Physical Execution Layer, Logic and Intelligence Layer, and Governance and Trust Layer — assigns distinct functional responsibilities to hardware abstraction, AI coordination, and compliance enforcement, enabling modular implementation and layer-independent evolution as the underlying technologies advance. The FRYX Protocol is designed as a genuine open standard, with governance implications extending to the standards-body structures necessary for its long-term maintenance and evolution.

The Discovery Economy model articulated in this paper describes the economic and scientific transformations that widespread DTIF adoption enables: the transition of autonomous laboratories from cost centers to IP-generating discovery infrastructure; the emergence of regional autonomous research hubs that democratize access to high-value scientific discovery; and the formation of global networks of autonomous research nodes that generate collective intelligence at scales unavailable to any individual laboratory. The realization of the Discovery Economy depends on governance conditions — open standards, transparent AI, global verification protocols, and equitable access frameworks — that are as important to the transformative outcome as the technical architecture itself.

The transition from human-driven experimentation to autonomous scientific systems will define the research infrastructure of the coming decades. The questions of how this transition is governed — who controls the infrastructure, how its outputs are verified, who benefits from its discoveries, and how its failures are attributed

and remediated — are among the most consequential policy questions of the scientific era that autonomous discovery is inaugurating. The DTIF provides a technical foundation upon which governance responses to these questions can be built: a concrete, operational architecture that makes abstract governance objectives — trust, transparency, sovereignty, and interoperability — measurable, enforceable, and improvable.

Future work will develop empirical validation studies of DTS parameter measurement methodologies across diverse experimental domains; formal specification of the FRYX instrument gateway registry and model translation ontology; case studies of DTIF implementation in pharmaceutical, materials science, and agricultural research contexts; and policy analysis of the international governance frameworks required for global verification protocol adoption. The DTIF specification presented in this paper represents an initial architectural proposal intended to stimulate the interdisciplinary engagement — across computer science, laboratory automation, science policy, ethics, and international law — that its full realization will require.

Appendices

Appendix A — SDL Maturity Assessment Template

The SDL Maturity Assessment Template provides a structured instrument for evaluating the autonomy level of an existing or proposed laboratory system against the SDL Maturity Model classification criteria. The template is designed to be administered by a qualified assessor with laboratory automation expertise in consultation with the research team responsible for system operation.

A.1 Assessment Scope and Method

Assessment is conducted across five primary dimensions, each evaluated on a subscale that maps to the SDL level boundaries. The five dimensions are:

(1) Hardware Automation Coverage — the proportion of experimental operations executable without direct human physical involvement;

(2) AI Integration Depth — the extent to which AI systems contribute to experimental design and decision-making;

(3) Human Oversight Architecture — the mechanisms through which human researchers monitor and intervene in autonomous operations;

(4) Data Provenance Infrastructure — the completeness and integrity of the data documentation and audit trail systems;

and (5) Inter-Agent Communication Capability — the degree to which the system can communicate and coordinate with external autonomous research agents.

Composite SDL level is determined by the lowest-scoring dimension, consistent with the principle that a system's autonomy classification reflects its weakest operational constraint.

A.2 Assessment Instrument Structure

For each assessment dimension, the evaluator responds to a set of structured questions with defined response options mapped to SDL level scores. Example questions include: 'What proportion of liquid handling operations are performed by robotic systems without human physical involvement? (0–25%: L0; 26–50%: L1; 51–75%: L2; 76–95%: L3; 96–100% with adaptive replanning: L4/L5).' The complete assessment instrument consists of forty-two questions distributed across the five dimensions, requiring approximately four hours to complete for a new laboratory system. Reassessment is recommended annually or following any significant modification to the system's hardware, software, or governance configuration. Assessment results should be documented in a standardized SDL Assessment Report and retained as part of the laboratory's quality management records.

A.3 SDL Level Certification

Institutions may seek SDL Level Certification from DTIF-recognized certification bodies, providing independent third-party validation of their self-assessment results. Certification is required for laboratories seeking authorization to participate in FRYX-connected global discovery networks at L4 or above, and is strongly recommended for laboratories operating at L3 in regulated research domains. Certification assessors must hold recognized qualifications in laboratory automation, AI systems evaluation, and quality management

systems, and must be independent of the institution under assessment. Certification is valid for two years, subject to interim audit requirements triggered by significant system modifications.

Appendix B — Digital Trust Score Calculation Guide

The DTS Calculation Guide provides operational specifications for computing the Digital Trust Score across diverse autonomous research contexts. The guide is intended for implementation by system integrators deploying the FRYX Digital Trust Engine and for use by regulatory bodies specifying DTS threshold requirements for their jurisdictions.

B.1 Verifiability Index Calculation

The Verifiability index is computed as a weighted average of three sub-component scores: Protocol Transparency Score (PTS), Instrument Calibration Integrity Score (ICIS), and Inter-Laboratory Replication Rate (ILRR). The recommended weighting for established laboratory systems with replication history is: $V = 0.30 \times PTS + 0.30 \times ICIS + 0.40 \times ILRR$. For newly deployed systems without replication history, the weighting reverts to $V = 0.50 \times PTS + 0.50 \times ICIS$ with a provisional flag applied to the composite DTS until at least five independent replication attempts have been completed. PTS is assessed by an automated protocol analysis tool that evaluates experimental protocols against the FRYX protocol schema compliance standard, returning a compliance fraction score. ICIS is computed from continuous instrument calibration monitoring data, calculated as the time-weighted fraction of measurement periods during which all active instruments were within calibration tolerance. ILRR is computed from the replication record maintained in the FRYX audit chain, calculated as the proportion of completed independent replication attempts that returned results within the defined tolerance bounds for the replicated experiment class.

B.2 Observability Index Calculation

The Observability index is computed as a weighted average of three sub-components: Inference Pathway Exposure Score (IPES), Feature Attribution Completeness Score (FACS), and Confidence Distribution Accessibility Score (CDAS). The recommended weighting is: $O = 0.40 \times IPES + 0.35 \times FACS + 0.25 \times CDAS$. IPES is assessed through an automated inference pathway analysis tool that traces the causal chain from model inputs to experimental recommendations and computes the proportion of decision nodes in that chain that are expressed in human-auditable form. FACS is computed by running the FRYX-standard attribution analysis pipeline — implementing SHAP value computation for all active AI models — and calculating the proportion of input features receiving attributable contribution scores above the minimum attribution threshold. CDAS is a binary or graduated score based on whether the agent exposes calibrated probability distributions (CDAS = 1.0 for fully calibrated distributions, 0.5 for uncalibrated confidence scores, 0.0 for point estimates only).

B.3 Traceability Constant Calculation

The Traceability constant is computed as: $Tr = PCC \times ACI \times TDC$, where PCC is the Provenance Chain Completeness fraction (0.0–1.0), ACI is the Audit Chain Integrity indicator (1.0 if cryptographically intact, 0.0 if any integrity violation is detected), and TDC is the Transformation Documentation Coverage fraction (0.0–1.0). The multiplicative structure of the Traceability formula reflects the non-substitutability of its components: a complete provenance chain that lacks cryptographic integrity protection, or a cryptographically secured chain with undocumented transformations, provides fundamentally weaker traceability guarantees than a system excelling on all three sub-components simultaneously. Any detected cryptographic integrity violation immediately reduces ACI to 0.0, thereby zeroing the Traceability constant regardless of PCC and TDC scores. This zero-tolerance policy for audit chain integrity violations reflects the critical importance of tamper-evidence in autonomous research governance.

B.4 Model Uncertainty Quantification

Model uncertainty (U) is computed as a composite of three sub-metrics: Ensemble Disagreement Index (EDI), Calibration Error (CE), and Out-of-Distribution Detection Rate (ODDR). The composite formula is: $U = EDI + CE - 0.5 \times ODDR$, where all sub-metrics are normalized to the [0, 1] range. The negative weighting of ODDR reflects the positive contribution of effective out-of-distribution detection to system reliability: a system that accurately identifies when it is operating outside its reliable domain and escalates appropriately is less dangerous than one that generates confident predictions in all input regimes. EDI is computed by comparing outputs from an ensemble of independently trained model instances on a held-out validation dataset. CE is computed using the Expected Calibration Error (ECE) metric on the same validation dataset. ODDR is computed as the proportion of out-of-distribution test cases (drawn from a curated OOD test suite specific to the research domain) correctly flagged by the agent's OOD detection mechanism.

Appendix C — FRYX Architecture Specification

The FRYX Architecture Specification provides technical reference documentation for the three-layer FRYX protocol stack. The specification is organized into three sub-sections corresponding to the Physical Execution Layer, Logic and Intelligence Layer, and Governance and Trust Layer.

C.1 Physical Execution Layer Specification

The Physical Execution Layer is specified through the FRYX Instrument Command Schema (FICS) — a JSON-Schema-compliant data format defining the structure of instrument commands, measurement requests, and data return specifications exchanged between the FRYX coordination infrastructure and laboratory hardware. FICS version 1.0 defines command classes for: liquid handling operations (aspiration, dispensing, mixing, transfer); thermal operations (incubation, heating, cooling, gradient cycling); imaging operations (brightfield, fluorescence, confocal, electron microscopy); spectroscopic measurements (UV-Vis, IR, NMR, mass spectrometry); and environmental monitoring (temperature, humidity, pressure, atmospheric gas composition). Each command class includes mandatory metadata fields — instrument identifier, calibration status token, authorization token, timestamp, and audit chain hash — that ensure all Physical Execution Layer events are captured in the provenance record. Extension namespaces for instrument classes not covered by the core specification are provided to accommodate specialized research domains without requiring core specification modifications.

C.2 Logic and Intelligence Layer Specification

The Logic and Intelligence Layer is specified through three interface standards: the FRYX Hypothesis Manifest Schema (FHMS), the FRYX Model Translation Ontology (FMTO), and the FRYX Multi-Agent Coordination Protocol (FMACP). FHMS defines the structured format for hypothesis specifications exchanged between the hypothesis generation engine and the multi-agent coordination system, including required fields for experimental design parameters, predicted outcome distributions, estimated information gain, resource requirements, and risk class designation. FMTO defines the shared scientific data ontology used by the universal model translator to map heterogeneous model input/output schemas to a common semantic representation, organized around top-level entity classes including Molecule, Material, Organism, Experimental Condition, Measurement, and Observation. FMACP defines the communication protocol for multi-agent task allocation and consensus, specifying message formats for task broadcast, bid proposal, task assignment, result submission, and consensus evaluation across FRYX-connected agent networks.

C.3 Governance and Trust Layer Specification

The Governance and Trust Layer is specified through two primary documents: the FRYX Digital Trust Engine API Specification (FDTE-API) and the FRYX Policy Language Specification (FPL). FDTE-API defines the programmatic interface through which FRYX-connected systems submit DTS computation requests, retrieve current agent trust scores, submit experiment authorization requests, and query audit chain records. The API is specified as a RESTful HTTP interface with OpenAPI 3.0 documentation, supporting both synchronous and asynchronous authorization request patterns to accommodate the latency requirements of different experimental domains. FPL defines the formal grammar and operational semantics of the FRYX policy specification language, enabling institutional governance authorities to express research governance rules in a machine-executable format that the Digital Trust Engine enforces at the point of experiment authorization. FPL is designed to accommodate the full diversity of national and institutional governance frameworks while

maintaining a shared semantic foundation that enables policy interoperability across FRYX-connected networks.

Appendix D — Cross-Agent Validation Protocol

The Cross-Agent Validation Protocol (CAVP) defines the structured communication procedures through which FRYX-compliant autonomous research agents verify each other's experimental claims prior to knowledge base integration. The protocol is designed to provide a distributed, scalable alternative to centralized peer review that can operate at the speed and scale of autonomous research production.

D.1 Validation Request Specification

A CAVP validation request is initiated by a research agent seeking validation of a specific experimental claim. The validation request must include: a structured specification of the claim to be validated in FRYX Hypothesis Manifest Schema format; the complete experimental record supporting the claim, including provenance chain documentation and DTS assessment at the time of original execution; an experimental protocol specification sufficient to enable independent replication; a risk class designation indicating the minimum DTS required of participating validation agents; and a consensus threshold specification indicating the proportion of successful independent validations required for the claim to achieve validated status. Validation requests are broadcast to the FRYX network through the FMACP multi-agent coordination channel, with validation agent selection governed by the Digital Trust Engine based on DTS scores, declared experimental capabilities, and conflict-of-interest assessments.

D.2 Validation Execution and Reporting

Accepting validation agents execute the specified experimental protocol independently, generating their own experimental records with full FRYX provenance documentation. Validation execution must be completed within the validity window specified in the original validation request — typically 30 to 90 days depending on the experimental domain and SDL level of the validating agent. Upon completion, validating agents submit a structured Validation Report that specifies: the validation outcome (Confirmed, Partial Confirmation, Not Confirmed, or Execution Failed); the experimental record supporting the validation outcome with full provenance documentation; a quantitative assessment of result agreement with the original claim, expressed as a normalized distance metric in the outcome space; and any anomalous observations encountered during validation execution that were not anticipated by the original experimental protocol.

D.3 Consensus Evaluation and Certification

The FRYX Digital Trust Engine aggregates individual Validation Reports submitted for a given claim, weighting each validating agent's outcome by its current DTS score. The consensus evaluation computes a weighted confirmation rate — the DTS-weighted proportion of validation outcomes classified as Confirmed or Partially Confirmed — and compares this rate against the consensus threshold specified in the original validation request. Claims achieving the required consensus threshold receive a FRYX Validated Claim Certificate, a cryptographically signed digital document specifying the claim content, the validation evidence base, the consensus rate achieved, and the certification timestamp. Validated Claim Certificates are the currency of knowledge integration in the DTIF framework: only certificate-bearing claims are eligible for integration into the shared scientific knowledge base and for submission to FRYX-compatible publication and regulatory workflows. Claims that fail to achieve consensus are returned to the originating agent with a detailed Validation Failure Report specifying the disagreement patterns observed across validation attempts, enabling targeted diagnostic investigation of the discrepancy.